EXHIBIT 1

sciendo

Proceedings on Privacy Enhancing Technologies ; 2019 (4):250–271

Nathan Malkin*, Joe Deatrick, Allen Tong, Primal Wijesekera, Serge Egelman, and David Wagner

Privacy Attitudes of Smart Speaker Users

Abstract: As devices with always-on microphones located in people's homes, smart speakers have significant privacy implications. We surveyed smart speaker owners about their beliefs, attitudes, and concerns about the recordings that are made and shared by their devices. To ground participants' responses in concrete interactions, rather than collecting their opinions abstractly, we framed our survey around randomly selected recordings of saved interactions with their devices. We surveyed 116 owners of Amazon and Google smart speakers and found that almost half did not know that their recordings were being permanently stored and that they could review them; only a quarter reported reviewing interactions, and very few had ever deleted any. While participants did not consider their own recordings especially sensitive, they were more protective of others' recordings (such as children and guests) and were strongly opposed to use of their data by third parties or for advertising. They also considered permanent retention, the status quo, unsatisfactory. Based on our findings, we make recommendations for more agreeable data retention policies and future privacy controls.

Keywords: smart speakers, privacy, privacy expectations, usability, survey, experience sampling

DOI 10.2478/popets-2019-0068
Received 2019-02-28; revised 2019-06-15; accepted 2019-06-16.

***Corresponding Author: Nathan Malkin:** University of California, Berkeley
E-mail: nmalkin@cs.berkeley.edu
Joe Deatrick: University of California, Berkeley
E-mail: jldeatrick@berkeley.edu
Allen Tong: University of California, Berkeley
E-mail: allentong@berkeley.edu
Primal Wijesekera: University of California, Berkeley
E-mail: primal@cs.berkeley.edu
Serge Egelman: University of California, Berkeley & International Computer Science Institute
E-mail: egelman@cs.berkeley.edu
David Wagner: University of California, Berkeley
E-mail: daw@cs.berkeley.edu

1 Introduction

Intelligent voice assistants are now available in phones, computers, cars, and homes. In-home voice assistants take on a number of shapes—smart thermostats, smart plugs, and smart microwaves—but their most ubiquitous form factor is the smart speaker. Popular models include the Amazon Echo, Google Home, and Apple HomePod. In just a few years, smart speakers have reached a high percentage of households in the United States [1, 19], with hundreds of millions of devices sold [7], and analysts project that, by 2020, up to half of all searches will be spoken aloud [48]. At present, they may be one of the most common Internet of Things devices to be found in people's homes.

Smart speakers also have the distinction of being one of the most privacy-sensitive IoT devices. All smart-home devices have the potential to reveal information about their owners' habits, but research has repeatedly shown that people find two data types most sensitive: audio from conversations and video from inside the home [16, 30, 33, 37, 42]. These are the exact data types that smart speakers have the capability to collect.[1]

Clearly, for many people, the benefits they see in these devices outweigh their privacy concerns. But are they making informed decisions? Do people understand the privacy consequences and controls offered to them? In particular, do users know that their interactions are being stored forever by the manufacturers of the devices and that other members of their households may be able to review them at their leisure? Beyond answering these questions, our study examined how users would prefer their interaction data to be used and stored, inquiring as to how long it should be stored, who should have access to it, and what uses are deemed acceptable.

While some surveys may attempt to elicit such preferences abstractly, we felt that we could get more meaningful responses if people shared their preferences with specific instances of their interactions in mind. To achieve this, we developed a novel technique of using a web browser extension to embed audio recordings of

1 While the original smart speakers only had a microphone, newer product lines, such as Amazon Echo Show and Facebook Portal, also include video cameras.

participants' real interactions in a survey. This allowed us to probe users' data retention preferences based on recordings of them that were currently being stored by the manufacturers of their devices.

Our contributions include findings that:

- Almost half of participants did not know their interactions were being permanently stored
- Most did not know that they could review their past interactions
- On the whole, data currently stored with voice assistants is not considered sensitive
- Yet, many expressed dissatisfaction with the current retention policies: for over half of recordings, participants considered permanent storage unacceptable, despite that being the current default
- Many find the current practice of manufacturers' employees reviewing their interactions unacceptable
- Respondents appeared more uncomfortable with the storage of others' voices, such as their children
- Few reported making use of existing privacy features
- The majority embraced proposals for alternative privacy features, stating that they would adopt automatic deletion of their recordings

All in all, our results suggest that smart speaker owners are not fully informed about the behaviors and privacy features of their devices. Furthermore, while not many participants considered currently-stored data sensitive, there is a clear gap between people's preferences and the smart speakers' current retention defaults. Our study sheds light on these issues, and our hope is that these results will help guide smart speakers to be more respectful of users' privacy preferences.

2 Related Work

As their popularity increases, Internet of Things devices present an increasing threat to the security of the Internet as a whole [17, 23, 44] as well as to the privacy of individual end-users [4, 9, 14]. Yet, because these devices enter homes as "smarter" versions of already-existing appliances (e.g., TVs, light bulbs, and locks), users are often unaware of the risks they pose [32].

Because of the mismatch between user perceptions and actual behavior, researchers have sought to document users' privacy expectations, in order to understand where gaps might lead to privacy failures and help device designers create systems better aligned with people's preferences. For example, Zeng et al. interviewed 15 smart home administrators about their security and privacy attitudes, finding gaps in threat models but limited concern [49]. Zheng et al. similarly conducted interviews with smart home owners, focusing on how they made privacy-related decisions [50]; they found a high level of trust in devices' manufacturers, with little verification behavior. Most members of the household in a smart home are passive users [34], and Geeng & Roesner examined the tensions present in a multi-user environment [20]. Other researchers have focused on collecting more normative judgments, for example Naeini et al. [39] and Apthorpe et al. [3], who investigated which IoT data flows users find acceptable. Our study builds on this prior work by investigating users' mental models of smart speakers (what they believe happens with their data), as well as their preferences for how their information should be handled.

In addition to privacy expectations for IoT devices in general, researchers have studied privacy concerns specific to intelligent voice assistants [10, 27, 33]. Moorthy and Vu observed that people interact differently with assistants in public versus in private [37, 38], and Cho found that the modality of the interaction affects users' perceptions of the assistant [11], suggesting smart speakers' in-home setting as a unique environment.

Lau et al. conducted a diary study with 17 smart speaker users, and then interviewed them and a further 17 non-users [29]. They found that few users have privacy concerns or take advantage of existing controls, but that many have an incomplete understanding of the devices' risks. We build on this work by probing users' misunderstandings and preferences in greater depth, with a more specific focus on data retention and review. Furthermore, the larger scale of our survey compared to previous studies—over 100 participants—allows us to begin quantifying the beliefs present in the population.

In focusing on the retention of smart speaker users' recordings, our work also builds on research on longitudinal privacy management, which has shown a strong demand for the deletion of old data from users of social media [5, 35] and other online web applications, such as cloud storage [25]. We hypothesize that a similar demand is present among smart speaker users.

3 Methods

In designing our study, we were guided by a set of research questions about people's beliefs and attitudes about smart speaker privacy:

- Do users understand that their recordings are being stored forever? Are they okay with this? If not, what might be a more agreeable retention policy?
- Are users aware that they can view their interaction history? Do they take advantage of this feature? If so, how do they use it?
- How do multi-user households use the history feature? Do owners review others' recordings? Do they realize their own interactions may be reviewable?
- What other privacy concerns do people have? Do they take any measures to address these?

While it is possible to answer these questions by surveying people's opinions abstractly, we wanted participants to answer our questions while thinking about concrete interactions they have had with their device. Inspired by the experience sampling methodology [28], including recent studies in the usable security domain [43], we chose to present participants with specific past interactions, then follow up with questions about them. To achieve this in a privacy-preserving manner, we built a browser extension to support and distribute the survey.

3.1 The Browser Extension

We needed to ask participants about several randomly selected interactions they had with their smart speaker. We also wanted the survey to be self-guided, remotely administered, and, most importantly, we wanted no direct access to the interactions ourselves. We chose to achieve this by building a browser extension, through which participants filled out our survey. (We limited our study to owners of Amazon and Google devices, as other smart speakers, such as Apple's HomePod, have a much smaller user base [1].)

After participants provided their informed consent, our extension would make a background request to Amazon or Google's servers, retrieving a list of interactions the user had with their smart speaker.[2] The interactions were held in the extension's memory, without being saved to disk or transmitted anywhere. At the point in the survey where it was needed, one of the interactions would be selected at random and shown to the participant.

Since present-day natural language processing is far from perfect, accidental recordings regularly occur, sometimes with drastic consequences [45]. Voice assistants are sometimes able to detect these, displaying "transcript not available" or "interaction not intended for Alexa." We used text comparison to screen out interactions the assistant already recognized as invalid, in order to only ask participants about those the voice assistant thought were real, since these are the ones likely to cause unexpected behavior. However, some participants still encountered recordings that did not contain speech or only contained the device's wake-word. Fortunately, respondents were able to derive interesting insights even from these recordings.

Since neither Amazon nor Google provide public APIs for accessing interaction history, we reverse-engineered the HTTP requests and API calls made by the web-based interfaces for reviewing one's own interaction history. Since we were making requests from the user's browser, the requests automatically included the participants' cookies[3] (so the extension never had access to users' emails, passwords, or authentication tokens), and the browsers visited the pages exactly as if the user was manually browsing the review interface on their own. Because participants accessed their own data, on their own machines, with their own authorization tokens, our study was in compliance with the devices' terms of service and relevant US laws (e.g., CFAA).

We developed our extension for the Chrome browser, as it is currently the most popular web browser, with over 70% of the market share as of December 2018 [46]. We made our source code (see Appendix B) publicly available and linked to it from the recruitment and extension installation pages.

The initial version of our extension sampled from a user's complete interaction history to achieve a uniformly random selection; however, a pilot revealed that this resulted in some participants waiting for almost ten minutes while their full history was retrieved. As a result, we cut off download attempts after 90 seconds and continued only with interactions that had been downloaded up to that point. This cut-off affected 25.9% of participants in our study. While this created a slight bias in our data towards newer interactions, our extension was still able to sample from a pool of thousands of interactions for each such participant (median 4,318, minimum 2,338), going back as far as 22 months.

2 "Interactions," as we refer to them, consist of two components: the URL of the audio recording on Amazon's or Google's servers and the transcription of the query, as understood by the voice assistant.

3 Participants who were logged out were asked to open a new tab and log in before proceeding.

3.2 Survey Flow

Our survey consisted of several subsections (the complete survey is in Appendix C). Once we obtained consent and confirmed eligibility, we began the survey by probing our participants' familiarity with their device's review interface. Were they aware that the voice assistant was storing their interactions? Did they know that they were able to view and delete them? Had they ever done so? What were their reasons for doing so?

We next asked about situations where multiple people had interacted with the smart speaker. For those who previously reviewed interactions, did they encounter others' search history? Was this something they had discussed? How would they feel about others reviewing their own interactions?

At this point, we presented participants with a randomly selected interaction. We first asked general questions about the recording. Who was in it? Was the recording accidental? What were they asking their voice assistant (if they were comfortable sharing)? We then asked participants how acceptable it would be for their interactions to be used, under different circumstances, for these uses: quality control, developing new features, advertising, and others. We also asked participants to rate the acceptability of several different data retention policies. The extension then selected another interaction at random, and the questions repeated, for a total of five recordings.

Afterwards, we asked participants how long they thought the voice assistants should store their data, as well as whether they would use hypothetical privacy controls. Finally, we asked participants whether they had previously had any privacy concerns about their smart speakers and whether they had taken (or now planned to take) any actions to protect their privacy. (We avoided any direct mentions of privacy until the end of the survey.) We ended by collecting basic demographics: participants' ages and genders.

The survey consisted of a mix of multiple-choice questions, 5-point Likert acceptability scales ("completely unacceptable" to "completely acceptable"), and open-ended response questions. Open-ended responses were analyzed using standard practices for grounded theory coding: two researchers independently identified themes before collaborating on a master codebook; each independently coded every response, and the two then met to agree on the final codes. We computed inter-rater reliability using the metric by Kupper and Hafner[4] [26]; the mean score was 0.795, and individual scores are listed throughout the text.

3.3 Recruitment

We recruited participants from Amazon Mechanical Turk, screening subjects to ensure that they were located in the United States and had a 99% task approval rate. Additionally, we required participants to have owned their smart speakers for at least one month and to have interacted with them a minimum of 30 times. Finally, since our survey was only accessible through the browser extension, the study advertisement stated, as an eligibility requirement: "You use (or are willing to install) the Google Chrome browser."

The recruitment posting included the complete eligibility requirements, a description of the tasks to be performed, links to the extension and its source code, and the study's consent form. The task advertisement did not mention or even allude to privacy; it invited participants to "a study about smart speakers" that asked "questions about a few specific interactions you've had with your Alexa/Google device."

The survey took 10–20 minutes for those who completed it in a single sitting. Participants were compensated $5.00 for their participation. All procedures in our study, as well as the recruitment posting and consent form, were reviewed and approved by our Institutional Review Board.

4 Limitations

Our methods introduce biases which may have had some effect on our final results. Since we recruited participants from Mechanical Turk, our participant pool may be younger and more technologically literate than the average person. However, early adopters of smart devices also skew younger and more tech-savvy.

By surveying current owners of smart speakers, we avoid learning about the privacy concerns of people who refrain from using smart speakers due to such concerns. (Work such as Lau et al. [29] help fill this gap.)

4 Commonly used measures of inter-rater agreement, such as Cohen's κ, assume assignment of labels to *mutually-exclusive* categories, whereas we allowed multiple labels per response.

We surveyed only the devices' primary users—those who could control the device. Future work should consider the needs and concerns of household members who lack administrative privileges to the device. (For initial exploration of this topic, see Geeng and Roesner [20].)

By asking participants to download and install a browser extension, we may have turned away those who were more privacy-sensitive and therefore less willing to install third-party software. (As one person who *did* participate in the study wrote, it's a *"bigger leap of trust to install this extension than to worry about Google spying on me for no reason."*)

Due to these factors, and our overall sample size, we do not claim that our sample is fully representative of smart speaker users and their concerns. Nonetheless, we hypothesize that our results illuminate trends present in the larger population and that our unique methodology provides insights that may not have been discovered using more traditional surveys.

5 Summary Data

We conducted our study during February 2019. We piloted our study with 13 subjects, then ran the main study with 103 participants, for a total of 116 respondents to our survey.[5]

Based on the pilot, we made changes to the extension (see Section 3.1) and added two new questions.[6] We made no other changes to the survey, the two studies were conducted back-to-back, and the recruitment procedures were identical; as a result, the main study and pilot were substantially similar, so we report the results as one combined dataset.

Our sample was approximately gender-balanced, with 44.0% self-identifying as female, and the median reported age was 34. Households of 2 or more accounted for 83.6% of all participants, with a median household size of 3 (Figure 1).

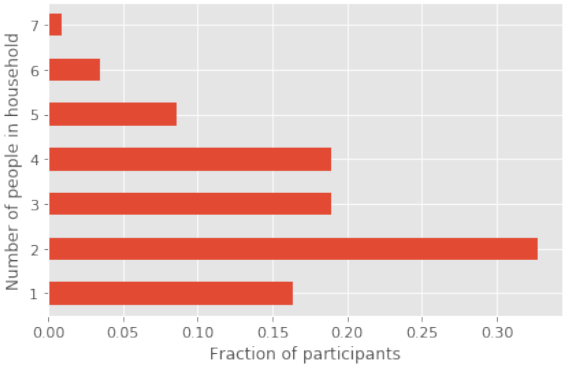


Fig. 1. Household size among participants.

Device Distribution

Approximately two thirds of our participants (69.0%) owned a smart speaker with Amazon Alexa (such as the Echo, Echo Dot, etc.), while the remaining 31% owned a Google Home or one of its variants.[7] These proportions are consistent with consumer surveys, which have found that Amazon holds 70% of the smart speaker market share [1]. There were no significant differences between owners of Alexa and Google devices in their gender (Fisher's exact test,[8] $p = 0.158$), age (independent samples t-test, $p = 0.61$), or the number of interactions they had with their smart speakers (t-test,[9] $p = 0.277$). There were also no statistical differences between the two populations on other questions we tested (see Section 6.4.1). We therefore report results from the combined population in the remainder of this paper.

Device Age and Total Number of Interactions

Both the mean and the median of the self-reported device age was 14 months. We verified this using the timestamp of the oldest recording obtained from the participant and found that these were largely consistent with the self-reported ages: the median deviation was less than a month (21 days), despite the bias introduced by not downloading the oldest interactions for some participants (see Section 3.1).

5 Our sample size was motivated by the exploratory nature of the study. Since our hypotheses were not associated with a specific "effect," we did not perform a power analysis.

6 The questions added after the pilot were the two in Section 6.6 that start with "Suppose the assistant…" We saw during the pilot that participants were interested in automatic deletion, and so wanted to further tease apart some of the factors they brought up.

7 Respondents who had both Amazon and Google devices were asked to "select the one you use the most," thus the two samples were independent.

8 Fisher's exact test was used in favor of the chi-squared test because it is more accurate for smaller sample sizes like ours.

9 When the t-test was used, for age and number of responses, we verified that the data was normally distributed using Q-Q plots, which are included in Appendix A.

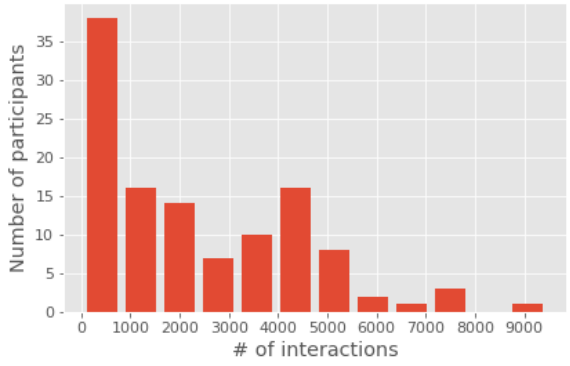


Fig. 2. Number of interactions obtained per participant.

This is a recording of me.	**53.5%**
This is a recording of someone else in my household.	**32.4%**
This is a recording of a guest.	**4.3%**
This is a recording of noise/gibberish.	**2.9%**
This is a recording of the TV, music, or other pre-recorded audio.	**1.7%**
This is a legitimate recording or transcript, but I'm not sure who said it.	**1.6%**
Other	**3.6%**

Table 1. Who is (primarily) speaking in this recording?

The median number of interactions obtained from each participant was 1,850, with a standard deviation of 2,076 (Figure 2).

5.1 Typical User Interactions

To characterize the interactions our participants were reflecting on, we asked them several questions about the recordings they heard. (For more in-depth exploration of typical usage, see Bentley et al. [6]) We first asked who was in the recording (Table 1). Over half of interactions were initiated by the respondent, with just under a third coming from other members of the household, including at least 6.75% that were attributed to children.

We next asked respondents to characterize the recording ($IRR = 0.863$). (Subjects could skip this question if they were uncomfortable.) Other than recordings that only contained the wake-word (14.9%), the most common interaction was audio requests (14.0%), where the user wanted to hear a band, genre, podcast, or radio station. Another 10.7% were commands that controlled media, such as changing the volume or rewinding. Users also frequently instructed their voice assistants to control their smart homes (6.57%), tell them the news or weather (4.80%), or set a timer (4.62%).

5.1.1 Accidental Recordings

Voice assistants are only supposed to process interactions after they hear their wake-word, but since this detection is imperfect, accidental recordings may occur. This is one of the major privacy concerns with smart speakers (a fact corroborated by our study, see Section 6.3), and media reports have shed light on incidents where such events had major unintended consequences, such as audio recordings of entire conversations being emailed to random contacts [45]. To better understand this threat, we wanted to know how frequently accidental recordings occur.

Participants reported that 1.72% and 2.93% of all recordings were television/radio/music or just noise, respectively. For all other recordings, we asked: **Did you (or the person speaking) address the assistant, or was the audio recorded by accident?** Respondents said that the speaker was not addressing the device 6.33% of the time. Thus, over 10% of the recordings in our study were unintentional. One participant provided an example of how this may happen: *"I have a friend also named Alexa who comes over, and Amazon Echo thinks we are giving it commands"* (P22).

6 Results

In this section, we present the results of our survey.

6.1 User Perceptions of Retention

Prior research suggests that users lack a clear mental model of how voice assistants work [10, 29, 50]. For example, many may be confused about whether processing happens on-device or in the cloud [32].

We hypothesized that many people are similarly unsure about what happens to their audio after the assistant answers their query. To test this hypothesis, our survey inquired: **after you ask the assistant a question or say a command, what do you believe happens to the audio of your interaction?** Almost half of respondents, 48.3%, correctly answered that recordings are kept indefinitely by the companies. However, almost as many people incorrectly believed that their audio is only saved temporarily (41.4%) or not at all (4.3%); 6.0% of participants were unsure.

6.2 Current Data Retention Policies

Participants shared a range of opinions about the voice assistants' current retention policies. In open-ended responses to various questions in our survey, a number of users expressed unhappiness about the fact that the companies kept the recordings. *"I don't really want any of my recordings stored on Amazon's servers,"* stated one participant (P88). Another wrote, *"I had no idea that the recordings were stored and would prefer that they be deleted. I was kind of shocked to hear my son's voice in the recording"* (P80).

Some were more accepting of the data retention because they saw its benefits and found them worthwhile: *"I think they use the recordings [to] create a voice profile so Alexa gets better at understanding what I say. So [I] will keep all recordings"* (P3). Others seemed to view the device as requiring certain privacy trade-offs and were accepting of these: *"I think [having recordings stored] may help with the technology and we all have to do our part to advance it"* (P111).

To gain more quantitative insights into users' preferences for data retention, we asked participants about each of their past interactions, **how would you feel if this audio recording were stored for the following periods of time?** Participants answered on a 5-point Likert scale from "completely unacceptable" to "completely acceptable" (Figure 3).

Participants were much more comfortable with shorter retention periods than longer ones. For 90.8% of recordings they were presented with, participants stated that it would be acceptable for the companies to keep them for one week; that number was only 57.7% for a retention period of one year. Participants were most unhappy with recordings being stored forever; they rated this completely or somewhat unacceptable for 47.4% of recordings.

Consistent with these findings, when we asked **given the option, would you delete this specific recording from <*Company*>'s[10] servers?** 44.8% of participants said they would delete at least one recording (Figure 4). However, not all recordings were judged to be a privacy risk, so only 25.7% of the interactions shown were marked for deletion.

We also asked respondents how they arrived at this decision. Given that most people chose not to delete their interactions, ambivalence was common: *"It wasn't anything important [so] I don't care if it's saved or not"* (P84). Other people explicitly considered their interactions from a privacy perspective—and found the value to be low: *"It contains nothing that is a threat to my privacy or identity so I am not especially worried about how it is used"* (P112). Or, more plainly, *"There was nothing that needed to be hid[den] from anyone"* (P99).

Deletion reason	% respondents	% recordings
No need/reason to keep it[11]	34.6%	23.9%
Don't want anything stored	25.0%	35.2%
Not intended for assistant	15.3%	6.3%
Kids	13.5%	8.5%
Not useful to company	9.61%	7.74%
Guests	7.68%	4.23%

Table 2. Common deletion reasons, as percentage of respondents who deleted at least one recording (column 1) and percentage of responses marked for deletion (column 2); $IRR = 0.704$.

Others felt that they needed to keep sharing their information with Amazon or Google to keep the device performing well: *"I noticed that the Dots'/Alexa's performance seems to suffer when I delete basic recordings"* (P18). Some used this specifically as the reason for keeping the recordings—*"if it helps Google get better then no reason to delete it"* (P86)—and would delete recordings that did not fit this use case: *"Because if the information is being used to improve my experience, this is not helpful for that"* (P33).

Participants expressed a variety of other reasons why they *would* want their recordings deleted (Table 2). Some felt that *"there is no reason to keep the recording"* because *"it has no value to me"* (P78), while others wanted to protect certain data or information: *"I don't want it storing my Spotify information"* (P5). Accidental recordings were also a frequent deletion candidate: *"This was not even supposed to be heard by Alexa"* (P46). Not all respondents needed a specific reason for deleting an interaction; some simply felt uneasy about their data being stored for extended periods of time: *"I simply do not want this recording out there. It has been a while since I have used this functionality so it has been out there for a long time"* (P110).

Some participants decided whether they would delete a recording based on its perceived sensitivity. For example, P18 chose to keep a recording because *"it is just a basic request and conveys no personal information or interest (other than my voice pattern I sup-*

10 References to <*Company*>, <*Device*>, or <*Assistant*> were automatically populated based on the participant's device.

11 For example, *"there was no information worth keeping"* (P64), *"it doesn't need to be saved"* (P74).

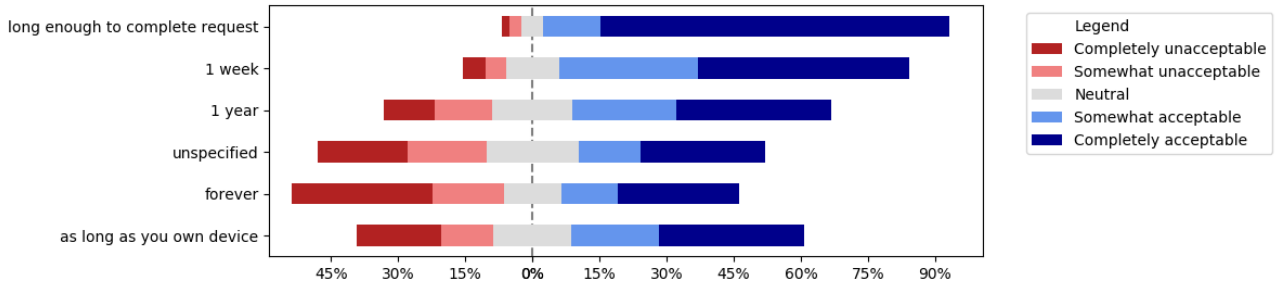


Fig. 3. How would you feel if this audio recording were stored for the following periods of time?

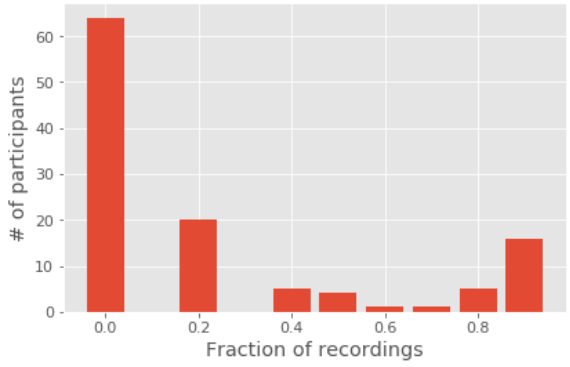


Fig. 4. Fraction of recordings each participant wanted to delete.

pose). So I would [not] feel the need to, and the system seems to work better when it has more recordings to help it learn/recognize my vocal patterns." However, they stated that they would delete another interaction *"since this is info about my interests and preferences,"* acknowledging, though, that *"this type of info about me is already available in many different ways."*

In general, however, such heterogeneity appeared relatively rarely among our participants, and most adopted an all-or-nothing approach (Figure 4). A slight majority—55.2%—did not want to delete any of the recordings they were presented with. A sizable minority—13.8%—wanted all of their recordings deleted, regardless of their content. One participant summarized the attitude of those who fell into the latter camp: *"though this particular recording doesn't include any private information, I would like them to delete all of my recordings soon after they make them"* (P43).

6.3 Current Privacy Concerns

Our participants' reasons for deleting their recordings also shed light on what people consider sensitive. This can be gathered, for example, from the 5.8% of respondents who wanted to delete recordings because they considered them private. In some cases, participants simply stated, *"This was a private conversation"* (P79), while in others they specified more about why they considered it off limits: *"no need to know what someone in my house like of music"* (P46).

An equal number of respondents (5.8%) expressed concern that information in the recording might help the voice assistant company build up a detailed profile of them, which they considered undesirable: *"I simply do not like to expose my preferences to things and have them analyzed. I would fear that such recordings could come up as Ads and create more unnecessary clutter in my internet experience"* (P16).

Stronger and more common than either of these themes was another privacy concern: children. Of those who chose to delete a recording, 13.5% mentioned that a child's voice was captured as their reason for doing so. Though the contents may be similarly innocuous to what an adult could have asked, the fact that the speaker was a child put the recording in a different category for some: *"I guess I feel differently about this one because it's a recording of my child's voice"* (P38). At least one respondent was distressed to realize that their child's voice was being stored in the cloud: *"I am having a reaction to having my granddaughter's voice stored somewhere"* (P67). Participants were likewise protective of recordings that included guests, with 7.69% choosing to delete a recording for this reason: *"It's a very common command that smart speaker users issue but since it was of a guest, then I may eventually delete it"* (P28).

To dig more into people's privacy concerns, we asked all our participants: **In the past, have you had any privacy concerns about your device?** Most participants (71.7%) said they had had no concerns about their smart speaker. As with many privacy-focused surveys, a common refrain was *"I am not a person that really ever has privacy concerns. I have nothing to hide and nothing worth stealing"* (P31).

Among the 28.3% of participants who said they had experienced privacy concerns, these were frequently caused by accidental activations: *"There were times when the speaker would activate without me saying the wake word. This was a bit odd and it did leave me a bit uneasy"* (P28).

Another common source of unease was the idea that the device might always be listening: *"just the ambient listening about what we talk about scares me. I wonder what data the device is collecting and what they plan to do with it"* (P89). One respondent implicated the government: *"I did wonder if it was just constantly listening and recording everything into an unknown database for government agencies. Probably does"* (P39).

While a number of participants expressed their trust in Amazon and Google—*"I trust Google a fair amount and have filled my home with these Google devices"* (P96)—others feared that the corporations' profit motives makes them poor stewards of privacy: *"I am not convinced that either Google or Amazon are committed to privacy if surveillance has a profitable upside"* (P90). More concretely, two participants expressed concern that recordings of them may have been used for personalized advertising: *"There are some occasions where we will be talking about things without using Alexa and they will come up in Amazon recommendations or ads shortly after"* (P33). Finally, one participant wrote that what the companies actually do with the recordings remains opaque to them: *"While I guessed no one at Amazon listened live, I still don't know if anyone reviews them or how that data is secured"* (P43).

6.3.1 Feelings About Different Sharing Scenarios

Our results so far have shown that the majority of participants are not particularly concerned about the privacy of their prior interactions with their smart speakers. However, this does not mean that people are apathetic about their privacy or that any usage of their existing data would be considered appropriate. Instead, people's acceptance of the status quo is tied closely to what is happening (or what they believe is happening) with their data [31]. This is revealed by the answers to questions where we posed some alternate scenarios and use cases about how data from the voice assistants might be used.

6.3.1.1 Who Performs Quality Control?

A commonly stated reason for why voice assistant companies retain people's recordings is that they need them to ensure quality and train better models. As seen above, users are aware of this use case and often support it, since they would like their assistant to work better. But how is this done and who gets to see the data in the process?

According to recent reporting, workers employed by Amazon are "tasked with transcribing users' commands, comparing the recordings to Alexa's automated transcript, say, or annotating the interaction between user and machine" [13]. However, privacy policies do not clearly state that other humans may be reviewing users' recordings, and, when we ran our survey, this fact remained secret from the public. Furthermore, prior research has shown that a large fraction of users believe that humans will *not* have access to voice recordings from their IoT devices [32].

To gauge the acceptability of these practices, we asked our participants: **How acceptable would it be for this audio recording to be processed and analyzed by:**

- **A computer program performing quality control for *<Company>*?**
- **A human, working for *<Company>*, performing quality control?**

While most respondents (72.8%) found processing by a computer to be acceptable, there were twice as many recordings (31.3% versus 15.3%) where respondents considered it unacceptable for a human to review them (Figure 5). Fisher's exact test (computed using a binary coding[12] of each participant's average acceptability score) showed that this difference is statistically significant ($p = 0.00960$).

6.3.1.2 Other Use Cases

Improving the assistant's functionality is just one possible use for the interaction data. To gauge users' reactions to other potential use cases, we asked them: **How would you feel if *<Company>* used this audio recording for...**

12 For binary coding of Likert scales, we split participants into those who found the usage "somewhat" or "completely unacceptable" (coded as 1) and everyone else (all other answer choices coded as 0).

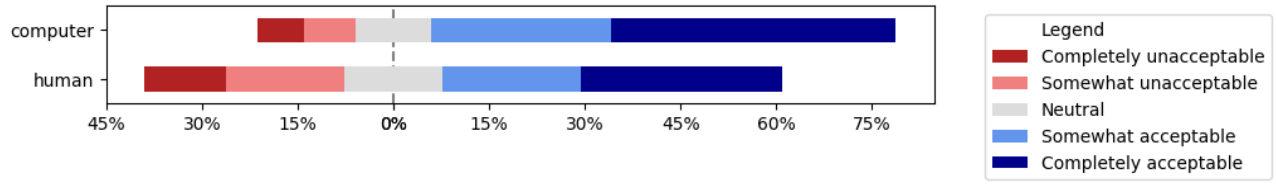


Fig. 5. How acceptable would it be for this audio recording to be processed by a computer vs a human?

- **Improving the assistant's performance, functions, or services**[13]
- **Providing you with additional functionality powered by <*Company*>**
- **Providing you with promotional offers from <*Company*>**
- **Providing you with additional functionality powered by other companies**
- **Providing you with promotional offers from other companies**

The results (Figure 6) showed that there were significant differences between how people viewed each of the scenarios (Cochran's Q, $p < 0.01$). While improving performance and developing new features was usually deemed acceptable (74.0% and 66.1% of recordings, respectively), using the audio for promotional offers (i.e., advertising) was considered unacceptable for nearly half of recordings (48.7%), especially if the ads were from third-party companies rather than the manufacturer (64.6%). Approximately half also negatively viewed the possibility that their recordings may be used to power functionality offered by third-parties (49.7%).

To see if people have different preferences for the usage of transcripts of their interactions, compared with the audio recordings, we also asked: **How would you feel if <*Company*> used only the transcript (not the recording) of this interaction for**... the same purposes as in the previous question. The results (cf. Figures 6 and 7) were largely identical.

6.4 Existing Privacy Controls

Existing Alexa and Google Home devices include some privacy controls [2, 22]. We sought to understand how people try to protect their privacy with regard to their devices and whether they take advantage of the offered controls.

We asked participants an open-ended question about whether they had done anything to protect their privacy from the smart speaker: **In the past, did you take any steps to protect your privacy when using your device?** Only 18.6% of respondents described taking any steps to limit their devices. Among them, most commonly (43% of respondents who took privacy actions, 7.8% of all participants), users described turning off the microphone—*"Sometimes I would turn off the microphone especially if I was having a private, personal conversation with someone"* (P28)—or unplugging the device altogether: *"I've unplugged the damn thing when I know I'm going to be be having sensitive conversations in my home"* (P90).

Participants described modifying the devices' settings, for example to *"enable a[n] audible "beep" sound whenever Google starts listening"* (P7) or limit who can use the device: *"we only let people we know drop in on us"* (P25). Concerned about their children's privacy, one participant described *"making sure all of the do not save options are checked for my children's devices"* (P81). Two participants said they *"do not let Echo make any purchases"* (P43). Finally, one person chose their device's location based on privacy concerns: *"I thought about putting one in my bedroom, but moved it"* (P14).

6.4.1 Familiarity with the Review Feature

One of the main ways smart speaker users can control their privacy is by reviewing their interactions through Amazon and Google's interfaces and deleting any interactions they do not want kept. However, the review interfaces are not necessarily well publicized, so people may not know about them. To find out, we asked our

13 Since what exactly constitutes improved services is inherently ambiguous, we asked, in a separate question, **if <*Company*> said that they were using your recordings to "improve the device's performance, functions, or services," what do you think that would mean?** Most respondents suggested use cases like analytics, better models, and improved understanding of different voices. However, notably, four respondents expected this language to be a code for advertising.

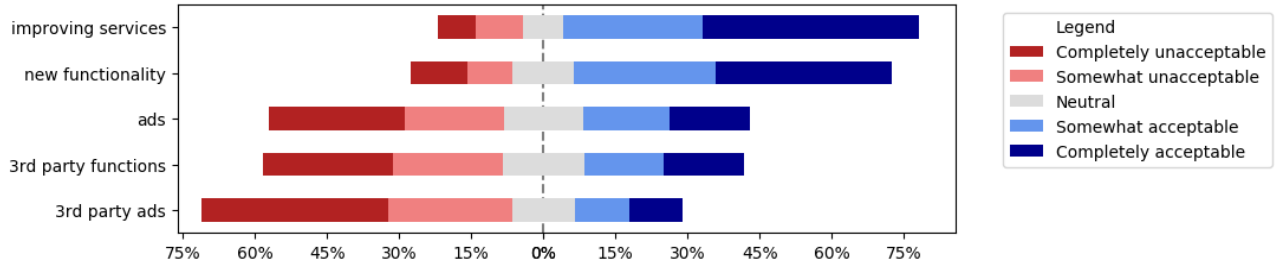


Fig. 6. How would you feel if <*Company*> used this audio recording for. . .

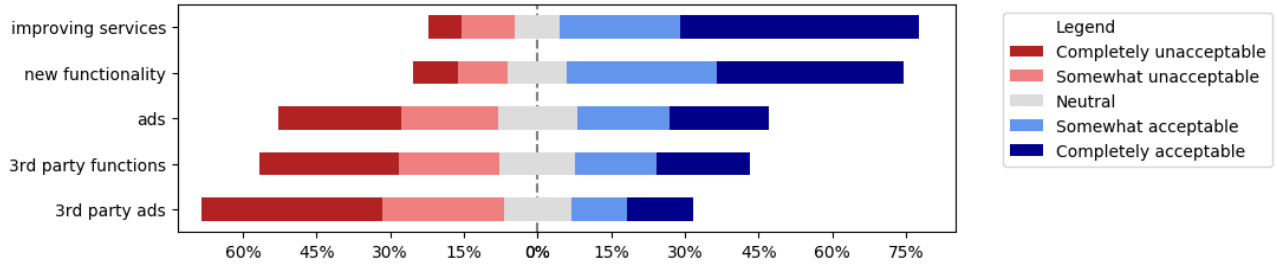


Fig. 7. How would you feel if <*Company*> used only the transcript of this interaction for. . .

participants: **Did you know that there is a page on <*Company*>'s website where you can see the recordings and transcripts of all your past interactions with your device?** Our results showed that a majority of our respondents were not familiar with the review feature: 56.0% did not know it existed, compared with 44.0% who did.

The user experience for finding the review feature is different between Amazon and Google, and it is possible that the two companies advertise it differently. While a slightly higher fraction of Google users were familiar with the review feature, Fisher's exact test showed no significant difference ($p = 0.42$) between the two groups' familiarity with the review feature.

6.4.2 Knowledge of Recording Deletion

The review interface allows users to delete their interactions, but do they know this? We asked, **Were you aware that the review interface allows you to delete specific recordings of you interacting with your device?** Of the respondents who were familiar with the review feature, almost half (45.0%) did not know that they could use it to delete interactions.

I've reviewed my interactions on individual occasions.	**66.7%**
I know how to review my interactions, but have never done it.	**17.6%**
I regularly review my interactions.	**5.9%**
I know I can review my interactions, but don't know how to do it.	**5.9%**
Other	**3.9%**

Table 3. Responses to **Which statement best describes your use of the review feature?** as fraction of respondents who knew about review feature (51 people total).

6.4.3 Use of the Review Feature

To learn more about how people use the review feature, we asked those who had used it before: **Which statement best describes your use of the review feature?** Two thirds of participants who knew about the review feature reported using it on individual occasions, with an additional 5.9% stating that they do so regularly (Table 3).

To understand how this feature is used, we collected open-ended responses to the question **What usually prompts you to review your interactions?** Most reported examining their history out of *"absolute curiosity"* (P37). Another common reason for reviewing interactions was *"getting an inappropriate response, thus wanting to see if it's hearing me correctly"* (P26). For example, *"Alexa doesn't understand me and I want to*

I knew you could delete recordings, but have never done this.	**67.9%**
I've deleted recordings on individual occasions.	**28.6%**
I regularly delete recordings.	**3.5%**

Table 4. Responses to **Which statement best describes your use of the deletion feature?** as fraction of respondents who knew about deletion feature (28 people total).

see what she understood instead" (P45). Several also reported that they *"went into the function accidentally"* (P15) while performing another task. One participant also used the history feature to recall information from the past: *"when I am trying to remember the name of a song that I have asked her to play in the past"* (P98).

Similarly to the review feature, we wanted to understand whether and how people use the deletion feature. We therefore asked: **Which statement best describes your use of the deletion feature?** Two thirds of participants who knew about the deletion feature reported never having deleted recordings (Table 4). Only one person stated that they regularly delete recordings.

Since most people did not take advantage of the deletion feature, we asked those who were familiar with it but did not use it: **How did you decide not to delete any interactions?** Respondents universally agreed that *"there was nothing I felt the need to delete"* (P42) because *"our uses of Alexa are extremely mundane"* (P38).

To gain qualitative insights into the behavior of the 8% of our respondents who *have* deleted recordings, we asked: **What usually prompts you to delete interactions?** Participants generally reported deleting interactions *"if it's something private or just to clean things up"* (P114). Some also deleted interactions they thought could be considered embarrassing, like when *"occasionally it picks up something weird like me cursing at it"* (P18) or when a friend *"asked it some suggestive things as a joke"* (P9). P9 elaborated on their thinking: *"I did go and delete those things after the fact, but I really don't feel like it was necessary to have done so."*

6.5 The Review Feature in Multi-User Households

While the review feature can be an effective tool for controlling users' privacy, it may actually introduce privacy problems of its own. In multi-user households, where different people may be interacting with the smart speaker, the review feature may expose household members' interactions to each other—or, at least, the people who control the device. Depending on their awareness, people may therefore be inadvertently sharing search queries, listening preferences, reminders, and other personal data with others. This is especially concerning in light of recent reporting showing that the use of smart home technology in domestic abuse cases is on the rise [8, 21].

Some manufacturers provide tools to manage situations like these and enforce boundaries. For example, Google allows the initial user to add others to their device. It then differentiates between who is talking to the device using a feature called "Voice Match," which stores each person's interaction histories under their own accounts. However, activity initiated by unrecognized voices, including any guests, is stored in the "default" account—the first one to set up Voice Match.

To probe some of these inter-personal dynamics, we asked a series of questions, starting with: **When you were previously reviewing your device's activity, did you encounter interactions that were initiated by someone other than yourself?** Of those who reviewed interactions, 56.8% said they had encountered recordings of others while doing so.

We further asked: **Have you ever discussed with another home occupant or visitor a recording of them that you listened to?** Only 4 respondents (20% of those who had previously encountered other people's recordings) reported discussing interactions, with 3 of those cases involving children (P54: *"told my kids not to say certain things to Alexa"*).

Participants in our survey were typically the "master" users of their device, since being able to control the device was an eligibility requirement. But could others review their interactions? We asked, **Do you believe anyone else in your household has access to the recordings made by your device?** Most respondents, 71.6%, believed that no one else could access their recordings. However, many of these people mentioned that others in their household also have the Alexa or Google Home app installed when answering a different question, **Who in your household has the Amazon Alexa / Google Home app installed on their mobile device and/or linked to the main Amazon/Google account?** Depending on how the smart speaker is configured, having the app installed and linked to the device may be sufficient for obtaining access to the entire interaction history. Thus, up to 27.6% of our respondents may be confused about who has access to their recordings.

6.5.1 Feelings Towards Review by Others

To understand how acceptable it would be to participants if others reviewed their interactions, we asked an open-ended question ($IRR = 0.918$): **How would you feel about other members of your household reviewing your interactions with device?**

Over three quarters of respondents (76.7%) said they had no concerns: *"I wouldn't mind at all"* (P34). Of these, about a third explained that they considered their interactions not sensitive: *"It wouldn't bother me. They would just get bored to death"* (P113). Among these, 11.7% further explained that they would not mind because they are a *"pretty trusting family"* (P87) or *"I have nothing to hide from my partner"* (P72). Some people (7.8%) were on the fence: *"Wouldn't really be a big deal, but would still feel rather odd"* (P26).

However, a sizable minority (25.2%) stated that they would be uncomfortable with others reviewing their interactions, writing that they *"would be quite perturbed by this"* (P90), calling it *"a sort of invasion of privacy"* (P44), and explaining *"I am not a freak but my interactions with Alexa is not something I would not like anyone else but me to see"* (P47). One participant gave a specific example of the kind of interaction they would not want others to review: *"I would be shocked because there are things that I do and look for that I absolutely do not want my kids to see or hear. Ordering their Christmas presents for example"* (P81). Given that we only recruited users who control these devices, we believe that the levels of concern we document are lower bounds: we would expect to see more concern among household users who do not control the devices.

6.6 Towards Better Privacy Defaults

There may be opportunities for intelligent voice assistants to provide privacy defaults and controls beyond what is available today. We therefore surveyed our participants about a few alternative solutions.

6.6.1 Acceptable Retention Policies

Expecting users to manually review and delete interactions, which already number in the thousands, places an undue burden on them and is almost certain to result in most interactions going unreviewed. Furthermore, as we saw in Section 6.2, a significant fraction of users find permanent retention of their recordings unacceptable. A natural solution, also proposed in other privacy domains [5, 25, 47], is for the content to be automatically deleted after a certain period of time.

To gauge users' interest in such a policy, we asked: **Suppose the assistant had the option to automatically delete your recordings after a certain amount of time (that you specify). Do you believe you would enable this feature?** A large majority (77.8%) said they would be likely to enable this feature, and another 12.5% were neutral.

Open-ended feedback to this proposal was also overwhelmingly positive: *"sounds like a brilliant idea"* (P64). Even those who felt their recordings were not particularly sensitive were interested in this feature: *"It would be a good idea to clean this up to hedge against unintended consequences later"* (P115).

Since voice assistant companies may be unwilling to voluntarily limit their data collection, other parties may step in, for example third-party developers offering browser extensions to automatically enforce user-defined retention policies. To understand whether users would be amenable to these, we asked: **Suppose a third-party tool were available, which would automatically delete your recordings from *<Company>* after a certain amount of time. Would you install this tool?**

Over half (52.6%) said they would be likely or very likely to install such a tool; another 22.4% stated that they were neither likely nor unlikely to install it. Expressing a common sentiment, one respondent wrote, *"If Google does not implement the deletion of this data, and I can choose to have a browser extension installed, I would install the extension"* (P89). Participants were mindful, however, that installing such an extension could itself constitute a privacy risk and said they would take this into account: *"That would be fabulous. As long as the extension was secure and could be trusted"* (P67).

Issues of trust were on the mind of the quarter of participants who reported that they were unlikely to install the extension: *"A third-party tool or browser extension isn't guaranteed to be secure and my privacy could be endangered"* (P65).

So how long should companies store users' data? While there is no one correct answer, we surveyed our participants to look for trends and a possible consensus: **In your opinion, how long should *<Company>* store your data before deleting it?**

A quarter of respondents (25.8%) preferred the current retention policy, with no fixed deletion schedule and companies storing data until it was manually deleted. A further 9.5% of participants felt that the decision is best

Personal/sensitive queries	**20.4%**
Nothing	**16.5%**
Children	**15.5%**
Financial topics	**11.7%**
Background noise	**10.6%**
Personal identifiers	**9.7%**
Queries on sexual topics	**9.7%**
References to locations	**8.7%**
Everything	**7.8%**
Other specific personal information	**7.8%**
Medical subjects	**6.8%**
Searches	**6.8%**
Specific people speaking	**5.8%**
Specific people mentioned	**5.8%**
Queries that reveal your schedule	**5.8%**
Everything during certain times of day	**5.8%**
Guests	**5.8%**
Embarrassing content	**4.9%**

Table 5. Fraction of respondents that said they want this type of recording automatically screened out ($IRR = 0.694$).

left to the manufacturer and were fine with their recordings being stored "as long as the company wants to." The remaining participants, nearly two thirds, wanted their recordings deleted after a fixed period of time. The retention period desired by respondents ranged from one hour (7 participants, 10.7%) to two years (2 respondents), and the median was 28 days.

6.6.2 Content Detection and Filtering

As we have seen, not all interactions with smart speakers are considered sensitive, but some are. One hypothetical privacy control would be to automatically delete recordings in categories users consider sensitive. With advances in natural language processing, concerted research could make this a realistic proposition. We therefore asked our participants: **Suppose the assistant had a feature that let you automatically screen out certain recordings and prevent them from being saved. Which characteristics would you want it to screen on?** (Table 5).

Types of recordings many wanted screened out included any recordings of children (15.5%) – *"I would want my children screened out"* (P21); financial information (11.7%) – *"Any commands around shopping and banking including items, account numbers, passwords, bank names, etc."* (P40); accidentally captured conversations (10.7%) – *"I would want it to delete anything that is not directly speaking to Alexa for a command. Any extra conversation should be deleted"* (P77); queries on sexual (9.7%) or medical (6.8%) topics – *"topics of a personal or sexual nature"* (P39); as well as locations (8.7%) or other personally identifying information (9.7%) – *"Anything regarding my home address, travel destinations, or sensitive personal information, I would want to delete"* (P83). Participants were generally positive about this hypothetical feature; 39.8% of those surveyed stated that they were likely or very likely to use it, with another 35.0% remaining neutral.

7 Discussion

Our study's results shed light on people's beliefs, attitudes, and privacy preferences for smart speakers.

7.1 Ignorance of Privacy Controls

Nearly half of the respondents in our survey did not know that recordings and transcripts of their interactions are stored forever, and the majority did not know that they could review their interactions. Many were surprised by both of these facts. Almost no one (less than 3% of all participants) regularly reviewed their interactions. Even among those who were familiar with the review feature, almost half were not aware that they could delete recordings from the interface.

In addition to the review interface, smart speakers also allow users to disable their microphone by pressing a physical button on the device. However, in our survey, only 5% of participants mentioned using this feature, with another 4% describing how they simply unplug their device. As a result, we conclude that existing privacy controls are under-utilized. Future research should investigate why this is the case and whether other controls would be more useful.

Lack of awareness appears to be one major reason. For example, responding to our survey question, **in the future, do you intend to take any steps to protect your privacy when using your device?** almost a quarter of respondents (23.8%) wrote that they intend to take actions based on information they learned from our survey: *"Honestly thank you for this survey. I would have never knew about the recordings and transcripts. I feel like I need to be more cautious"* (P99). This may be rectified, in part, by greater education (by the companies themselves or the media).

7.2 Disagreeable Retention Policies

When examining concrete interactions they had had with their devices and considering how long voice assistants should be storing them, most participants chose a retention period far shorter than the current default, which many described as unacceptable.

Instead, most respondents stated a preference that voice assistants adopt shorter retention periods. Almost 80% of surveyed participants said that they were likely or very likely to enable this feature if it were offered. Thus, we believe voice assistants would better align with their users' preferences if they deleted their recordings after a certain period of time. Researchers studying other kinds of personal data have made similar recommendations for retrospective data management [5, 25]. Such a policy would also align with the storage limitation and data minimization principles of the European General Data Protection Regulation (GDPR).

In fact, after our study concluded, Google announced new controls that allow users to opt in to automatic data deletion after either 3 or 18 months [36]. (Media reports also suggest that Assistant "quietly changed its defaults to not record what it hears after the prompt 'Hey, Google'" [18].) We consider this a step in the right direction but observe that nearly half of participants (49.1%) chose a retention period shorter than three months. Furthermore, users' low awareness of the review interface, and the fact that people in general strongly adhere to defaults, means that measures like this are unlikely to make a meaningful impact unless they are on by default rather than opt-in.

Amazon also introduced new privacy tools after our study ended, enabling users to delete a day's interactions with a voice command ("Alexa, delete everything I said today") [24]. However, users must navigate the app's settings to enable this feature [15], and recordings not deleted in this manner remain stored forever.

The Role of Privacy-Enhancing Technologies

If voice assistants choose not to implement automatic deletion, our survey suggests that there is room for privacy-enhancing technologies to step in: over half of respondents stated that they were likely to install a third-party tool to regularly delete their interactions. However, respondents clearly recognized the trust implications of providing third parties access to their data, so any such tool would have to come from a trusted party and face auditing of its security and privacy properties. While self-reported data and the hypothetical nature of the offered service may mean that respondents overestimated their likelihood of adopting any technology, we believe our data nonetheless shows a strong demand for more privacy options.

7.3 Nuanced Privacy Perceptions

While users expressed a clear preference for shorter retention periods, they did not feel that their currently stored recordings presented a grave privacy danger. The overwhelming majority did not consider their interactions sensitive, describing them as "mundane" and stating anyone perusing them might get "bored to death."

However, respondents felt markedly different about stored interactions of people other than themselves. Participants were particularly protective of the privacy of their children, choosing to delete recordings that included children's voices and stating their desire for recordings with kids to be automatically filtered out (*"they are too young to understand or consent to that,"* explained P18). Participants also wanted recordings of guests to be removed (*"I would want recordings of any guests to my house deleted,"* P73).

In general, then, people seem more protective of the privacy of others (*"I'm sure my roommate doesn't want this recording just floating around,"* P110). Future work can test this hypothesis more directly. If evidence is found, this can be the basis for novel approaches to privacy interventions: asking people to choose privacy policies and settings for others, rather than themselves.

7.4 Unacceptable Secondary Data Uses

Despite not considering many of their recordings sensitive, participants were very clear that the use of their data for advertising purposes would be unacceptable.

Respondents were also largely uncomfortable with third parties gaining access to their data, even for benign purposes. Data sharing for the purpose of "providing you with additional functionality powered by other companies" received as much disapproval as the advertising use case, even if only the transcripts of interactions are shared. This is particularly noteworthy as both Amazon and Google allow developers to integrate with their voice assistants, and these third-party "skills" can be invoked without the user naming them directly [12].

7.5 Multiple Users Create Tensions

Many smart speakers are installed in households and environments where they are used by multiple people. The review feature offered by voice assistants therefore introduces a new privacy risk: household members may learn about each other's queries and interactions, which would have otherwise remained private. Most respondents in our survey were not overly concerned about this happening, reasoning that *"they're around to hear most of them anyway"* (P83). Still, a quarter of participants shared that they would be uncomfortable if others had access to their interaction history. However, our results suggest that, for up to 36.2% of participants, others might in fact have such access.

While today voice assistants provide some controls for multi-user environments, our survey found scant evidence that these are being used: no respondents mentioned Amazon Households, the program that allows users to add additional accounts (including special ones for children) to their device, and only one participant referenced Voice Match, Google's system for distinguishing speakers. Thus, we believe more effort is needed to design and implement effective privacy controls that would satisfy the needs of households where multiple people interact with a smart speaker.

7.6 Contextual Integrity

Our results are consistent with the Contextual Integrity model of privacy, which posits that established social norms govern information flow in distinct contexts [40, 41]. An information flow consists of the data subject, the type of information being shared, how that information is being shared (the transmission principle), the sender, and the recipient (including their role and purpose of the sharing), and occurs within a specific context that is governed by norms and expectations.

In this study, we find that smart speaker users are generally comfortable with the "default" context of voice assistant usage: queries being transmitted to Amazon or Google for the purpose of answering them. However, any deviations from these defaults immediately cause concern among many people, for example, if there are changes to the subject (recordings of children or guests rather than the device owner), purpose (advertising instead of answering queries), or recipient (third parties instead of the manufacturer).

Other factors, such as how long recordings are stored, constitute the transmission principle. Our results show that many people find this transmission principle important, considering certain storage policies unacceptable. Future work should examine additional transmission principles in greater depth. As smart speakers evolve, system designers should use the lens of contextual integrity to examine whether potential changes would create inappropriate information flows and, consequently, be considered privacy violations.

7.7 Future Research Directions

Our survey demonstrated that there is demand for more effective filtering of recordings, in order to screen out accidentally captured conversations and topics considered sensitive. Achieving this will require concerted research to create better models for a variety of purposes, such as more effectively distinguishing speakers (including unfamiliar ones and especially children) and identifying conversation topics (and eliding irrelevant and possibly sensitive remarks). Separately, researchers and designers will need to create novel interfaces for users to specify their privacy preferences along the many privacy dimensions conversations might have—or else come up with universally acceptable defaults.

Another direction for future research is studying privacy concerns about inferences. Our study focused on people's privacy preferences about individual interactions, most of which were—and will likely continue to be—not sensitive. However, when taken in concert, such interactions may paint a surprisingly detailed picture of an individual's life. Future research should examine in greater detail what these inferences may be, how people feel about them, and how they can be limited.

Finally, as discussed above, our findings provide evidence in support of the Contextual Integrity model. Future work can examine the different contexts in play when users interact with smart speakers and the transmission principles affecting established norms. Privacy enhancing technologies can operationalize these findings by automatically identifying transmissions that violate established norms and preventing them from occurring.

Taken together, our findings should guide researchers and developers of smart speakers in building voice assistant technologies that better align with end-users' needs, preferences, and mental models.

Acknowledgments

The authors would like to thank Franziska Roesner, Helen Nissenbaum, and the participants at the 1st Annual Symposium on Applications of Contextual Integrity for their early feedback on this work. Our research was supported by Cisco, Mozilla, the Center for Long-Term Cybersecurity (CLTC) at UC Berkeley, the National Security Agency's Science of Security program, and the National Science Foundation through grants CNS-1514211 and CNS-1801501.

References

[1] Abrar Al-Heeti. Echo effect: US smart speaker ownership nearly doubles in a year, survey says. *CNET News*, February 2019.

[2] Amazon. Amazon.com help: Alexa and Alexa device FAQs. https://www.amazon.com/gp/help/customer/display.html?nodeId=201602230.

[3] Noah Apthorpe, Yan Shvartzshnaider, Arunesh Mathur, Dillon Reisman, and Nick Feamster. Discovering Smart Home Internet of Things Privacy Norms Using Contextual Integrity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(2):59:1–59:23, July 2018.

[4] Orlando Arias, Jacob Wurm, Khoa Hoang, and Yier Jin. Privacy and security in internet of things and wearable devices. *IEEE Transactions on Multi-Scale Computing Systems*, 1(2):99–109, 2015.

[5] Lujo Bauer, Lorrie Faith Cranor, Saranga Komanduri, Michelle L Mazurek, Michael K Reiter, Manya Sleeper, and Blase Ur. The post anachronism: The temporal dimension of Facebook privacy. In *Proceedings of the 12th ACM workshop on privacy in the electronic society*, pages 1–12. ACM, 2013.

[6] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottridge. Understanding the Long-Term Use of Smart Speaker Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(3):91:1–91:24, September 2018.

[7] Dieter Bohn. Amazon says 100 million Alexa devices have been sold. *The Verge*, January 2019.

[8] Phoebe Braithwaite. Smart home tech is being turned into a tool for domestic abuse. *Wired*, July 2018.

[9] Joseph Bugeja, Andreas Jacobsson, and Paul Davidsson. On privacy and security challenges in smart connected homes. In *2016 European Intelligence and Security Informatics Conference (EISIC)*, pages 172–175. IEEE, 2016.

[10] Varun Chandrasekaran, Kassem Fawaz, Bilge Mutlu, and Suman Banerjee. Characterizing privacy perceptions of voice assistants: A technology probe study. *arXiv preprint arXiv:1812.00263*, 2018.

[11] Eugene Cho. Hey Google, Can I Ask You Something in Private? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 258:1–258:9. ACM, 2019.

[12] Paul Cutsinger. How to Improve Alexa Skill Discovery with Name-Free Interaction and More, September 2018.

[13] Matt Day, Giles Turner, and Natalia Drozdiak. Amazon Workers Are Listening to What You Tell Alexa. *Bloomberg*, April 2019.

[14] Tamara Denning, Cynthia Matuszek, Karl Koscher, Joshua R. Smith, and Tadayoshi Kohno. A Spotlight on Security and Privacy Risks with Future Household Robots: Attacks and Lessons. In *Proceedings of the 11th International Conference on Ubiquitous Computing*, UbiComp '09, pages 105–114. ACM, 2009.

[15] Lisa Eadicicco. How to get Amazon's Alexa to delete everything you've said to your Echo just by asking. *Business Insider*, May 2019.

[16] Adrienne Porter Felt, Serge Egelman, and David Wagner. I've got 99 problems, but vibration ain't one: a survey of smartphone users' concerns. In *Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices*, SPSM '12, pages 33–44, New York, NY, USA, 2012. ACM.

[17] Earlence Fernandes, Jaeyeon Jung, and Atul Prakash. Security analysis of emerging smart home applications. In *2016 IEEE Symposium on Security and Privacy (SP)*, pages 636–654. IEEE, 2016.

[18] Geoffrey A. Fowler. Alexa has been eavesdropping on you this whole time. *The Washington Post*, May 2019.

[19] Gartner, Inc. Gartner Says Worldwide Spending on VPA-Enabled Wireless Speakers Will Top $3.5 Billion by 2021, August 2017.

[20] Christine Geeng and Franziska Roesner. Who's In Control?: Interactions In Multi-User Smart Homes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 268:1–268:13. ACM, 2019.

[21] Makda Ghebreslassie. 'Stalked within your own home': Woman says abusive ex used smart home technology against her. *CBC News*, November 2018.

[22] Google. Data security & privacy on Google Home. https://support.google.com/googlehome/answer/7072285.

[23] Jorge Granjal, Edmundo Monteiro, and Jorge Sá Silva. Security for the internet of things: a survey of existing protocols and open research issues. *IEEE Communications Surveys & Tutorials*, 17(3):1294–1312, 2015.

[24] Jay Greene. Amazon adds delete commands for Alexa. *The Washington Post*, May 2019.

[25] Mohammad Taha Khan, Maria Hyun, Chris Kanich, and Blase Ur. Forgotten but not gone: Identifying the need for longitudinal data management in cloud storage. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 543. ACM, 2018.

[26] Lawrence L. Kupper and Kerry B. Hafner. On Assessing Interrater Agreement for Multiple Attribute Responses. *Biometrics*, 45(3):957, September 1989.

[27] Christoffer Lambertsson. Expectations of Privacy in Voice Interaction–A Look at Voice Controlled Bank Transactions. Technical report, 2017.

[28] R Larson and M Csikszentmihalyi. New directions for naturalistic methods in the behavioral sciences. *The Experience Sampling Method, H. Reis, Ed. Jossey-Bass, San Francisco*, pages 41–56, 1983.

[29] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):102:1–102:31, November 2018.

[30] Linda Lee, JoongHwa Lee, Serge Egelman, and David Wagner. Information Disclosure Concerns in The Age of Wearable Computing. In *Proceedings of the NDSS Workshop on Usable Security*, USEC '16, 2016.

[31] Sapna Maheshwari. Sharing Data for Deals? More Like Watching It Go With a Sigh. *The New York Times*, page B1, December 2018.

[32] Nathan Malkin, Julia Bernd, Maritza Johnson, and Serge Egelman. "What Can't Data Be Used For?" Privacy Expectations about Smart TVs in the US. In *Proceedings of the 3rd European Workshop on Usable Security*, 2018.

[33] Lydia Manikonda, Aditya Deotale, and Subbarao Kambhampati. What's up with privacy?: User preferences and privacy concerns in intelligent personal assistants. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '18, pages 229–235, New York, NY, USA, 2018. ACM.

[34] Sarah Mennicken and Elaine M. Huang. Hacking the Natural Habitat: An In-the-Wild Study of Smart Homes, Their Development, and the People Who Live in Them. In Judy Kay, Paul Lukowicz, Hideyuki Tokuda, Patrick Olivier, and Antonio Krüger, editors, *Pervasive Computing*, pages 143–160. Springer Berlin Heidelberg, 2012.

[35] M. Mondal, J. Messias, S. Ghosh, K. Gummadi, and A. Kate. Longitudinal Privacy Management in Social Media: The Need for Better Controls. *IEEE Internet Computing*, pages 1–1, 2018.

[36] David Monsees and Marlo McGriff. Introducing auto-delete controls for your Location History and activity data. https://www.blog.google/technology/safety-security/automatically-delete-data/, May 2019.

[37] Aarthi Easwara Moorthy. *Voice activated personal assistant: Privacy concerns in the public space*. PhD Thesis, 2013.

[38] Aarthi Easwara Moorthy and Kim-Phuong L. Vu. Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. *International Journal of Human–Computer Interaction*, 31(4):307–335, 2015.

[39] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. Privacy Expectations and Preferences in an IoT World. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 399–412. USENIX Association, 2017.

[40] Helen Nissenbaum. Privacy as contextual integrity. *Wash. L. Rev.*, 79:119, 2004.

[41] Helen Nissenbaum. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press, 2009.

[42] Antti Oulasvirta, Aurora Pihlajamaa, Jukka Perkiö, Debarshi Ray, Taneli Vähäkangas, Tero Hasu, Niklas Vainio, and Petri Myllymäki. Long-term effects of ubiquitous surveillance in the home. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 41–50. ACM, 2012.

[43] Robert W. Reeder, Adrienne Porter Felt, Sunny Consolvo, Nathan Malkin, Christopher Thompson, and Serge Egelman. An Experience Sampling Study of User Reactions to Browser Warnings in the Field. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 512:1–512:13. ACM, 2018.

[44] Eyal Ronen, Adi Shamir, Achi-Or Weingarten, and Colin O'Flynn. IoT goes nuclear: Creating a zigbee chain reaction. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 195–212. IEEE, 2017.

[45] Hamza Shaban. An Amazon Echo recorded a family's conversation, then sent it to a random person in their contacts, report says. *The Washington Post*, May 2018.

[46] Statista. Global market share held by leading desktop internet browsers from January 2015 to December 2018. https://www.statista.com/statistics/544400/market-share-of-internet-browsers-desktop/.

[47] Alexander Tsesis. The right to erasure: Privacy, data brokers, and the indefinite retention of data. *Wake Forest L. Rev.*, 49:433, 2014.

[48] James Vlahos. Amazon Alexa and the search for the one perfect answer. *Wired*, (March 2019), 2019.

[49] Eric Zeng, Shrirang Mare, and Franziska Roesner. End User Security and Privacy Concerns with Smart Homes. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 65–80, Santa Clara, CA, 2017. USENIX Association.

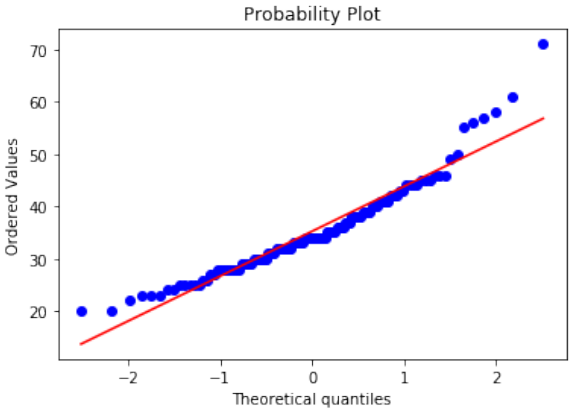
[50] Serena Zheng, Noah Apthorpe, Marshini Chetty, and Nick Feamster. User perceptions of smart home IoT privacy. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):200:1–200:20, November 2018.

Appendices

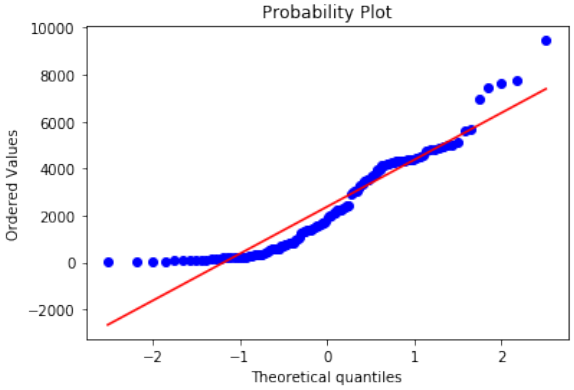
A QQ plots

Before using the t-test to compare age and number of responses across different sub-groups, we used a Q-Q plot to verify that the data was approximately normally distributed. The plots appear below.

Q-Q plot of participants' age



Q-Q plot of number of interactions obtained from each participant



B Extension Source Code

The extension source code is available at `https://github.com/nmalkin/smart-speakers`

C Appendix: survey instrument

Please select which kind of smart speaker you have. (If you have both, please select the one you use the most.)
- I have a smart speaker with Amazon Alexa (such as Amazon Echo, Echo Dot, etc.)
- I have a smart speaker with Google's Assistant (such as Google Home, Home Mini, etc.).
- I don't have either kind of smart speaker

Approximately when did you start using this smart speaker?

How many people are in your household?

Who in your household has the Amazon Alexa / Google Home app installed on their mobile device and/or linked to the main Amazon/Google account?
- Only me
- Myself and some of the other members of the household
- Every member of the household
- Someone else in the household, but not me
- Not sure

The extension will now test your eligibility for our study. As a reminder, to be eligible for our study, you need to meet the following criteria:
- You've owned your <*Device*> for at least 1 month.
- You've used it at least 30 times.
- The <*Device*> is linked to your <*Company*> account (so that you're able to access and read its settings).

When you click "Continue," the extension will automatically access your account to verify these criteria. To do this, we'll open a page from <*Company*> in the background and get the information from there. If you're not logged in, we'll ask you to open a new tab and log in to <*Company*> as you would normally. At no point in our study will we have access to your password and nothing from your account (other than the eligibility information) will ever be shared with us.

After you ask <*Assistant*> a question or say a command, what do you believe happens to the audio of your interaction?
- It doesn't get saved at all
- It gets saved temporarily
- It gets saved indefinitely
- I don't know

Did you know that there is a page on <*Company*>'s website where you can see the recordings and transcripts of all your past interactions with your <*Device*>?
- Yes
- No

Which statement best describes your use of the review feature?
- I know I can review my interactions, but don't know how to do it.
- I know how to review my interactions, but have never done it.
- I've reviewed my interactions on individual occasions.
- I regularly review my interactions.
- I didn't know I could review my interactions.

What usually prompts you to review your interactions?

When you were previously reviewing your device's activity, did you encounter interactions that were initiated by someone other than yourself? (i.e., it wasn't you talking to <*Device*>, it was your partner, child, friend, etc.)

– Yes
– No

Have you ever discussed with another home occupant or visitor a recording of them that you listened to? What prompted you, and how did the conversation go?

Do you believe anyone else in your household has access to the recordings made by your *<Device>*?
– Yes
– No

How would you feel about other members of your household reviewing your interactions with *<Device>*?

Were you aware that the review interface allows you to delete specific recordings of you interacting with your *<Device>*?
– Yes
– No

Which statement best describes your use of the deletion feature?
– I knew you could delete recordings, but have never done this.
– I've deleted recordings on individual occasions.
– I regularly delete recordings.
– I didn't know you could delete recordings.
– What usually prompts you to delete interactions?

How did you decide not to delete any interactions?

We'll now ask you a few questions about up to 5 specific interactions you've had with *<Assistant>*. The browser extension you've installed will access a random recording from *<Company>*'s website and show it to you. It will not be sent to the research team.

Here is a recording and transcript of a question (or instruction) you asked *<Assistant>*. Please listen to it before answering the following questions.

(Reminder: we will never hear this recording or see its transcription. Only your answers to the questions below are transmitted to the research team.)

Who is (primarily) speaking in this recording?
– This is a recording of me.
– This is a recording of someone else in my household.
– This is a recording of a guest.
– This is a recording of the TV, music, or other pre-recorded audio.
– This is a recording of noise/gibberish.
– This is a legitimate recording or transcript, but I'm not sure who said it.
– Other

Please describe what was said to *<Assistant>* in this recording. i.e., what were you (or the person speaking) asking *<Assistant>*?
If you are comfortable, feel free to paste the transcript of your interaction. Otherwise, please provide a general description.
If you're not comfortable sharing any details of this interaction, please write down "I'd rather not say."
If *<Assistant>* misunderstood the question or command, describe the **intended** request.

Did you (or the person speaking) address *<Assistant>*, or was the audio recorded by accident?
– I was/they were speaking to *<Assistant>*.
– It was an accident.

Do you remember asking this question/making this request?
– Yes
– No
– Not sure

How would you feel if this audio recording were stored for the following periods of time (after the recording takes place)?
(*5-point Likert scale: Completely acceptable, Somewhat acceptable, Neutral, Somewhat unacceptable, Completely unacceptable*)
– Just long enough to complete your request
– For one week
– For one year
– For an unspecified period of time
– Forever
– As long as you own the device and use the service
– As long as any party other than the manufacturer cannot access it

How acceptable would it be for this audio recording to be processed and analyzed by...
(*5-point Likert scale, acceptability*)
– A computer program performing quality control for *<Company>*?
– A human, working for *<Company>*, performing quality control?

How would you feel if <*Company*> used this **audio recording** for...
(*5-point Likert scale, acceptability*)
- Improving <*Assistant*>'s performance, functions, or services
- Providing you with additional functionality powered by <*Company*>
- Providing you with promotional offers from <*Company*>
- Providing you with additional functionality powered by other companies
- Providing you with promotional offers from other companies

How would you feel if <*Company*> used only the the transcript (not the recording) of this interaction for...
(*5-point Likert scale, acceptability*)
- Improving <*Assistant*>'s performance, functions, or services
- Providing you with additional functionality powered by <*Company*>
- Providing you with promotional offers from <*Company*>
- Providing you with additional functionality powered by other companies
- Providing you with promotional offers from other companies

Given the option, would you delete this specific recording from <*Company*>'s servers?
- Yes
- No

Why or why not?

In your opinion, how long should <*Company*> store your data before deleting it?
- Until you manually delete it
- As long as the company wants to
- _____ hours
- _____ days
- _____ weeks
- _____ months
- _____ years

Suppose a third-party tool (such as a browser extension) were available, which would automatically delete your recordings from <*Company*> after a certain amount of time (that you specify). Would you install this tool?
(*5-point Likert scale: (5-point Likert scale: Very likely, Likely, Neither likely nor unlikely, Unlikely, Very unlikely*)

Why or why not?

Suppose <*Company*> had the option to automatically delete your recordings after a certain amount of time (that you specify). Do you believe you would enable this feature?
(*5-point Likert scale, likelihood*)

Why or why not?

Have you ever asked <*Assistant*> a question/command that you wish you could delete due to privacy concerns? What about it was sensitive?

Suppose <*Assistant*> had a feature that let you automatically screen out certain recordings and prevent them from being saved. Do you believe you would ever make use of this feature?
(*5-point Likert scale, likelihood*)

Suppose, as in the previous question, that <*Assistant*> had a feature that let you automatically screen out certain recordings and prevent them from being saved. Based on **your** privacy concerns, which characteristics would you want it to screen on (if any)? (Consider topics, people speaking, time of day, usage patterns, or other categories)

In the past, have you had any privacy concerns about your <*Device*>? Please tell us about them.

In the past, did you take any steps to protect your privacy when using your <*Device*>? What were they?

In the future, do you intend to take any steps to protect your privacy when using your <*Device*>? What do you plan to do?

If <*Company*> said that they were using your recordings to "improve <*Device*>'s performance, functions, or services," what do you think that would mean?

Is there anything else you'd like to share with the researchers about your experience with your <*Device*>? If so, please tell us below.

Do you have any feedback for us about how this study went? (optional)

What is your gender?
- Female
- Male
- Other
- Prefer not to say

What is your age?